

# An Improved Link Model for Window Flow Control and Its Application to FAST TCP

Krister Jacobsson, Lachlan L. H. Andrew, *Senior Member, IEEE*, Ao (Kevin) Tang, *Member, IEEE*, Steven H. Low, *Fellow, IEEE*, and Håkan Hjalmarsson

**Abstract**—This paper presents a link model which captures the queue dynamics in response to a change in a transmission control protocol (TCP) source's congestion window. By considering both self-clocking and the link integrator effect, the model generalizes existing models and is shown to be more accurate by both open loop and closed loop packet level simulations. It reduces to the known static link model when flows' round trip delays are identical, and approximates the standard integrator link model when there is significant cross traffic. We apply this model to the stability analysis of fast active queue management scalable TCP (FAST TCP) including its filter dynamics. Under this model, the FAST control law is linearly stable for a single bottleneck link with an arbitrary distribution of round trip delays. This result resolves the notable discrepancy between empirical observations and previous theoretical predictions. The analysis highlights the critical role of self-clocking in TCP stability, and the proof technique is new and less conservative than existing ones.

**Index Terms**—Acknowledgments (ACKs), fast active queue management scalable transmission control protocol (FAST TCP).

## I. INTRODUCTION

WITHIN the field of network congestion control [23], [30], one line of work of fundamental interest is the dynamics of congestion control protocols, such as transmission control protocol (TCP). These use feedback to adapt sources' sending rates to the network resources. Control theory provides a suitable mathematical framework for the analysis and synthesis of such systems [10]. Conversely, the emerging field of networked control systems [11], in which control messages are transmitted over networks, requires accurate knowledge of the delay and jitter introduced by the network. Both of these require sufficiently accurate yet tractable dynamical models of network elements.

Initial TCP performance analysis relied heavily on simulation, but since the late 90s [17], *network fluid flow models* have

dominated. These models abstract away packet level details by viewing a flow of packets as a smooth fluid, and explicitly model feedback from the network to senders. These models have allowed control theory to be applied to analyze both local stability [6], [13], [16], [20], [21], [24]–[26], [35] and global stability [1], [8], [9], [12], [27], [38], [41] of congestion control algorithms. Stability is important to ensure fluctuations due to stochastically varying cross traffic are damped, and the network operates in a favorable region of the state space. Unstable protocols cause small fluctuations in cross traffic to produce large fluctuations in queue lengths, which reduce throughput and increase jitter, which interferes with interactive services such as voice-over-IP.

Despite their success, existing fluid models discard too much packet level information, and can produce qualitatively inaccurate results. To allow model based design of control protocols, recent results [40] suggest that it is sometimes crucial to consider more packet level phenomena.

Current TCPs are window based; each sender controls a window size, which is an upper bound on the number of packets that have been sent but not acknowledged. The rate of transmission is controlled or “clocked” by the received acknowledgments (ACKs): a new packet is transmitted only when an ACK is received, thereby keeping the number of outstanding packets equal to the window size. Thus, sources control the amount of data they inject into the network rather than the rate of doing so. Intuitively, this “volume control” is safer in terms of stability than “rate control”.

Until recently, most network stability analyses have modeled sources as controlling their data rates explicitly; the rate of change of the queueing delay is then proportional to the difference between the aggregate incoming traffic and link capacity. We refer to this as an integrator link model. The rate each window-based source inputs to the links is assumed to be the window size divided by the round trip delay, a relation which holds in equilibrium but not during transients. Typically, these models predict that the system is stable when round trip delays do not exceed some upper bound. This is in line with the intuition that increased feedback delay may have a destabilizing effect on a closed loop system.

The integrator link model however completely ignores self-clocking. Self-clocking has a major impact; in extreme cases, it causes the increase in queueing delay to be proportional to the change in window, rather than its integral. Self-clocking is captured by the “static” link model [36], [37] which, for a single bottleneck, reduces to assuming that the number of packets in the queue is exactly the sum of the windows minus the bandwidth-delay product. Using the static link model, it has been theoretically shown that fast active queue management scalable

Manuscript received May 08, 2007; revised November 06, 2007. Current version published March 11, 2009. This work was supported by the National Science Foundation under Grant 0303620, the WAN-in-Lab Project, also supported by the Caltech Lee Center for Advanced Networking, and Cisco, and by the Swedish Research Council under Contract 621-2004-4695. Recommended by Associate Editor I. Paschalidis.

K. Jacobsson and S. H. Low are with the California Institute of Technology, Pasadena, CA 91125 USA.

L. L. H. Andrew is with the CAIA, Swinburne University of Technology, Victoria 3122, Australia (e-mail: lachlan.andrew@gmail.com).

A. Tang is with Cornell University, Ithaca, NY 14853 USA.

H. Hjalmarsson is with the Royal Institute of Technology (KTH), Stockholm SE-100 44, Sweden.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TAC.2009.2012986

TCP (FAST TCP) flows [39] are always stable for the case of homogeneous sources [8], [36], [37]. As the static link model fails to hold when round trip delays are heterogeneous, we need a better model to study the general case.

Both the integrator and the static link models are incomplete in the sense that each only emphasizes one side of the story; the former lags the true dynamics, while the latter leads the true dynamics yielding optimistic predictions (See Examples 1, 2, 3 and 5 following). After introducing notation and background material, we will show in Section II that a natural combination of these two models leads to a more accurate one. Open loop experiments validate this joint link model and reveal its underlying intuition.

The new joint link model predicts significantly different queue trajectories from existing models on time scales comparable to a round trip time. Recently, many protocols have been proposed which use delay to respond on this time scale; for example, TCP Vegas [5], Africa [18], Fusion [19], Adaptive Reno [28], Compound TCP [31] and FAST [39]. Accurate analysis of all of these protocols requires the use of a model similar to the joint link model. This is illustrated in Section III, which investigates the stability of FAST TCP. Using the joint link model, we prove that FAST running over a single bottleneck link is stable for any heterogeneous delays, and hence resolve the discrepancy between previous experimental results and existing theoretical predictions. Closed loop experiments are also reported where accurate predictions on the stability region are obtained and verified with packet level simulations<sup>1</sup>.

## II. MODEL AND NOTATION

To capture the self-clocking effect in window flow control, we avoid working directly with the sources' sending rates. Instead, we use the sources' window sizes and the bottleneck queue size to represent the state of the closed loop system.

### A. Preliminaries

Consider  $N$  window-based TCP sources sending over a bottleneck link with capacity  $c$ . Let  $w_n(t)$  denote the congestion window of source  $n$  at time  $t$ ,  $n \in \{1, \dots, N\}$ . Let a packet that is sent by source  $n$  at time  $t$  appear at the bottleneck queue at time  $t + \tau_n^f$ . This *forward delay*  $\tau_n^f$  models the amount of time it takes to travel from source  $n$  to the link, and it accounts for the constant forward latency but not queueing delays. The *backward delay*  $\tau_n^b(t)$  is the time from when a packet arrives at the link to when the corresponding acknowledgment is received at source  $n$ , including the time-dependent queueing delay at the bottleneck queue. The *round trip delay* seen by source  $n$ , denoted  $\tau_n(t) = \tau_n^f + \tau_n^b(t)$ , is the elapsed time between when a packet is sent and when the corresponding acknowledgment is received. The *latency* of source  $n$ , denoted  $d_n$ , is defined as the round trip delay when the bottleneck queue is empty.

<sup>1</sup>To the best of our knowledge, the current status of research on congestion control protocols can provide quantitative results on equilibrium, while for dynamics, most works focus on qualitative study and have not been able to compare predictions with packet level simulations quantitatively.

The queueing delay of the bottleneck link is denoted by  $p(t)$ , and  $c > 0$  is the capacity of the link. The queueing delay observed by the  $n$ th source at time  $t$  is  $q_n(t)$ ; it relates to the queue delay by  $q(t) = p(\tilde{t})$ , where  $\tilde{t}$  solves  $t = \tilde{t} + \tau_n^b(\tilde{t})$ .

The bottleneck link may also carry non-window-based traffic such as User Datagram Protocol (UDP) traffic. Let  $x_c(t) \in [0, c]$  be the rate at which such cross-traffic is sent over the link, leaving  $c - x_c(t)$  for window based sources.

Whenever a time argument of a variable is omitted it represents its equilibrium value; for example,  $p$  is the equilibrium value of  $p(t)$ . In discrete time,  $w[k]$  denotes  $w(t_k)$ .

### B. Link Models

As described in Section I, previous work differs in how the dynamic map between the window sizes and the buffer size is modeled. These models reflect behavior on different timescales.

1) *Long Time Scales, The Integrator Model:* A queue simply integrates the excess instantaneous input rate at the link minus the capacity. Most existing literature on window based congestion control [2], [4], [13], [21]–[23] makes the assumption, valid on long time scales, that the sending rate is the window size divided by the round trip delay. This gives

**[Integrator link model]**

$$\dot{p}(t) = \frac{1}{c} \left( \sum_{n=1}^N \frac{w_n(t - \tau_n^f)}{d_n + p(t)} + x_c(t) - c \right). \quad (1)$$

2) *Short Time Scales, The Static Model:* The model (1) does not take into account the “self-clocking” of window based schemes, where the sending rate is regulated by the rate of the received ACKs. This is dominant on short time scales.

Consider  $N$  flows sending over a fully utilized bottleneck link with constant window sizes, and consider the response to a change in window size by a system initially in equilibrium.

The rate of packets flowing out from the link, and hence the sum of the rates of the received ACKs of the  $N$  sources, equals the capacity of the link. Increasing the window injects additional packets into the network, increasing the instantaneous rate for a very short time. Since the link is fully utilized (input traffic equals the capacity), these extra packets are appended to the queue. Thus, the queue's immediate response to a window increase is just a proportional increase in the queueing delay  $p$  one forward delay  $\tau_n^f$  after the window  $w_n$  is changed (i.e., when the new packets reach the bottleneck).

In the idealized case without cross traffic and with all flows having an equal round trip time (RTT), the sum of the rates of the ACKs again equals the capacity of the link and there is no further transient (sources' sending rates are auto-regulated by their individual ACK rates). This yields the “static” link model proposed in [36], [39] and implicitly used in [29]. The relation between the window size and the buffer size is then described by the following algebraic relation:<sup>2</sup>

<sup>2</sup>The original model was presented in discrete time for multiple bottlenecks, here we use its continuous time version used in, for example, [8] and we consider a single bottleneck.

**[Static link model]**

$$\sum_{n=1}^N \frac{w_n(t - \tau_n^f)}{d_n + p(t)} = c - x_c(t). \quad (2)$$

3) *Proposed Joint Link Model:* The immediate but short-lived response of the static model (2) contrasts to the smooth transient of the integrator link model (1), and both neglect important aspects of actual links. We now describe a more accurate “joint” link model, which captures effects on both timescales. This model is derived in [14] from a detailed analysis of the packet level system, a simplified version of which is in Appendix A.

Consider again the short term effect of a window change, this time in the presence of non-window-based cross traffic such as UDP.

There is again a step increase in queueing delay  $p$  proportional to the increase in window, but this time sources can affect their ACK rates over time intervals greater than one RTT. By sending a burst of packets back-to-back, a flow can achieve short term rate through the bottleneck equal to its capacity. This results in a burst of ACKs at that same rate, which in turn will increase the queue input rate one RTT later and so on, causing cascaded bursts of decreasing magnitude. The queue size integrates all of these short bursts of rate.

A similar effect occurs when there is no cross traffic but the heterogeneity among sources’ RTTs is significant. Individual flows operate on their individual RTT time scales, and it takes one RTT before a queue change affects the queue input rate. Thus, from the perspective of flows with small round trip times, flows with larger RTTs can be considered as non-responsive cross traffic, and the system is hence transient in this case also.

Tracking these bursts, spaced apart by different intervals for each flow, would yield an infinite-dimensional model. Instead, the joint link model captures the initial burst, and then models the subsequent “reflected” bursts as an increase in overall rate, averaged over the whole RTT, yielding an integrator similar to (1). From Appendix A, this gives

**[Joint link model]**

$$\dot{p}(t) = \frac{1}{c} \left( \sum_{n=1}^N \left( \frac{w_n(t - \tau_n^f)}{d_n + p(t)} + \dot{w}_n(t - \tau_n^f) \right) + x_c(t) - c \right) \quad (3)$$

which can be seen as a superposition of (1) and (2). The derivative term  $\dot{w}(t - \tau_n^f)$  models the immediate proportional change in the queue size due to a window change. Note that it is the window size and its corresponding time derivative *only* that have delayed variable arguments, which furthermore are identical to be the forward delay [14]. A similar model was also implicitly used for flow control stability analysis in [3].

Linearizing (3) by a standard first order Taylor expansion around the equilibrium defined by  $\sum_{n=1}^N w_n/(d_n + p) + x_c = c$ , gives the Laplace domain transfer function for flow  $n$  as

$$G_{pw_n}(s) = \frac{\mathcal{L}[p(t)]}{\mathcal{L}[w_n(t)]} = \frac{\frac{1}{c} \left( s + \frac{1}{d_n + p} \right) e^{-s\tau_n^f}}{s + \frac{1}{c} \sum_{i=1}^N w_i/(d_i + p)^2}. \quad (4)$$

For the case with homogeneous delay  $d_n = d$  and no cross traffic, applying the equilibrium identity  $\sum_{n=1}^N w_n/(d + p) = c$  shows that the pole and zero of (4) cancel, whence the map is a pure delay, scaled by  $1/c$ . Thus the model correctly reduces to (2) in this case. Even though (2) was described as the “short time scale” response, it is also accurate for gradual changes in  $w_n$  in this case, because there is no change in the ACK rates to cause a follow-on response.

Similarly, note that adding cross traffic for this scenario slows down the pole, since the sum in the identity then becomes less than the capacity  $c$ , while the zero remains fixed [14]. When cross traffic dominates, the pole dominates the zero, thus making the joint link model (3) approach the integrator link model (1). Note that with heterogeneity among the sources, depending on the configuration, the pole may be faster *or* slower than the zero. Finally, note that (4) is open loop stable, as expected due to self-clocking.

**C. Open Loop Validation Experiments**

The accuracy of the joint link model (3) as well as its similarities and differences with the integrator link model (1) and the static link model (2) are illustrated in the following open loop examples. Further validation examples of the joint link model can be found in [14] and a closed loop experiment will be reported in Section III-E. These simplistic scenarios have been chosen to demonstrate the relation between the static, integrator and joint link models.

The models (1), (2) and (3) are compared with packet level data generated by NS-2. In each experiment we consider 20 window based flows with static windows, sending over a single bottleneck link. Non-bottleneck links provide configurable forward and backward delays. The window sizes are initially set to the same constant size and are *not* updated dynamically, i.e., there is no dynamic feedback except for self-clocking. The system is started in equilibrium and perturbed at time  $t = 10$  s by a 10% step change in the first source’s congestion window. In all experiments a packet size of 1040 bytes is used.

1) *Example 1: Homogeneous Sources:* All 20 window based flows share the same latency  $d_n = 200$  ms and the bottleneck link capacity is  $c = 100$  Mbit/s. The scenario may represent parallel downloading of multiple files between two end users. The window size is  $w_n = 125$  packets. Source 1, which is subject to the window change, has a forward delay of  $\tau_1^f = 100$  ms. The solid gray line in Fig. 1 shows the queue size when the system is simulated in NS-2. The black dashed, dotted and solid lines show the integrator link model (1), the static link model (2) and the joint link model (3) respectively. The fit of the static and joint link models is excellent (neglecting packet level “noise”); they are identical in this scenario. This suggests that the true dynamics in this case is indeed a pure delay. Also observe that the integrator link model lags the NS-2 simulation. Note that it takes 100 ms before the window change affects the queue, as predicted by the models.

2) *Example 2: A Cross Traffic Scenario:* The scenario for Fig. 2 is as in Example 1 but with bottleneck link capacity  $c = 500$  Mbit/s which is also shared by 400 Mbit/s of UDP traffic. In this case the dynamics are clearly distinguishable; it takes over four seconds (twenty round trip times) before the

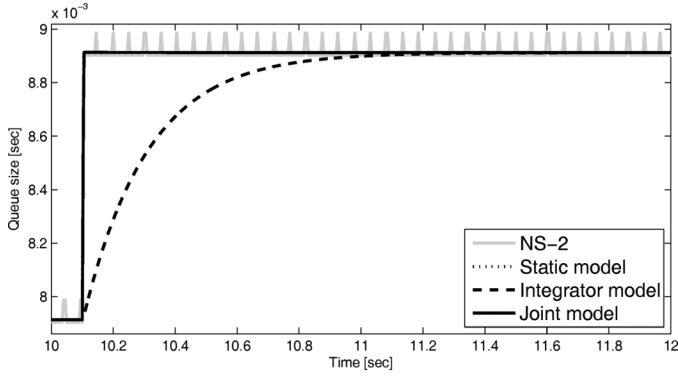


Fig. 1. Homogeneous sources. Both the joint link model and the static link model agree with the NS-2 simulation, but the integrator link model lags significantly.

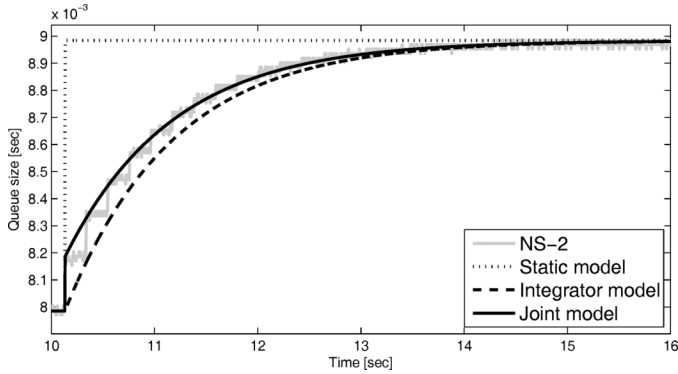


Fig. 2. Cross traffic. Both the joint and integrator link models agree with the NS-2 simulation, but the static link model leads significantly.

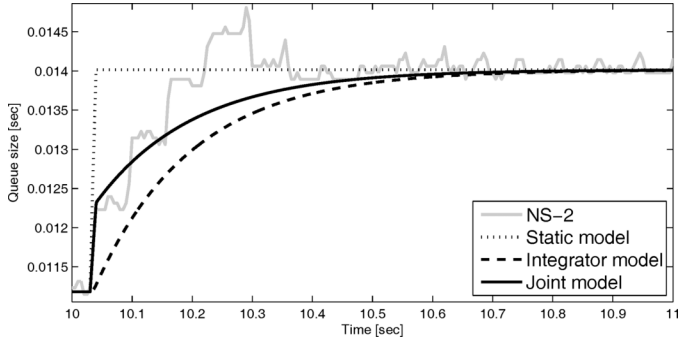


Fig. 3. Heterogeneous sources. The joint link model captures both the initial jump and the protracted rise; the integrator link model misses the former, while the static link model misses the latter.

queue settles again. The static link model is too rapid in this case, as expected, while the other two are both accurate. The joint link model captures the rapid initial rise in queue size, and initially tracks the upper envelope of the staircase simulation results, while the integrator link model tracks the lower envelope. From 1 s after the transient, the joint link model tracks the mean while the integrator link model lags slightly. As predicted, the joint model resembles the integrator in this case.

3) *Example 3: Heterogeneous Sources:* The scenario for Fig. 3 is as in Example 1 but the first source has round trip delay  $d_1 = 50$  ms, distributed such that  $\tau_1^f = 25$  ms, while  $d_n = 250$  ms for the remaining 19 sources. All sources have  $w_n = 135$  packets.

In this case there is a more pronounced initial increase in the queue followed by a transient phase which dies out after about 250 ms. This corresponds to the time until the self-clocking of the high delay sources adjust to the new conditions. The steps spaced by  $d_1 = 50$  ms in the NS-2 results show the response each time a new burst of ACKs arrives one RTT after the previous burst; capturing this sub-RTT bustiness requires a more detailed model.

In summary, these three examples demonstrate from different perspectives that while the integrator link model may lag and the static link model can lead the true dynamics significantly, the joint link model (3) succeeds in modeling the two main system characteristic of “self-clocking”, namely the short term proportional change, and the long term integrating effect that are present in the system.

### III. APPLICATION: STABILITY OF FAST TCP

To demonstrate the application and tractability of the joint link model, we will apply it to the stability analysis of FAST TCP [39]. FAST is a high speed TCP variant that uses delay as its main control signal. So far, all experiments with FAST have operated at a stable equilibrium regardless of how big the round trip delays are. This is at odds with the prediction of the integrator model in [37]; detailed reasons for this are discussed in Appendix B. This section will show that under the joint link model (3), FAST is indeed locally<sup>3</sup> stable for a single bottleneck link with the default step size.

Unlike most existing work, the analysis culminating in our main result, Theorem 3, uses a detailed model of FAST, which includes filters corresponding to the RTT estimation and the sampling processes. For the parameter settings used by the implementation of FAST, these filters noticeably improve stability.

#### A. Window and Estimator Model of FAST TCP

The sending rate of FAST TCP is implicitly adjusted via the congestion window mechanism. Each sender  $n$  updates its window size in discrete time according to

$$w_n[k+1] = (1-\gamma_n)w_n[k] + \gamma_n \frac{d_n}{d_n + \hat{q}_n[k]} w_n[k] + \gamma_n \alpha_n. \quad (5)$$

This update is performed once per RTT, with sampling time  $h_k$ . The parameter  $\alpha_n \in \mathbb{Z}^+$  is the number of packets that a FAST source tries to keep buffered in the network. It determines the equilibrium fairness, queueing and sensitivity to timing jitter. Despite recent progress [34], setting  $\alpha$  is an open problem. The gain parameter  $\gamma_n$  affects the protocol’s response rate [39].

The buffer occupancies are treated as continuous time dynamical processes, with quantization treated as noise. The aggregate queueing delay  $q[k]$  can be approximated by subtracting the latency  $d_n$  from the measured RTT. However, this gives a noisy measurement of the “true” queueing delay, which is therefore *estimated* by the source. We do not consider the problem of estimating  $d_n$ .

<sup>3</sup>Although only local stability has been established so far, all simulations satisfying the local stability conditions have reached the unique equilibrium, suggesting that the domain of attraction may indeed be the whole space. These preliminary linear results are useful to guide parameter selection.

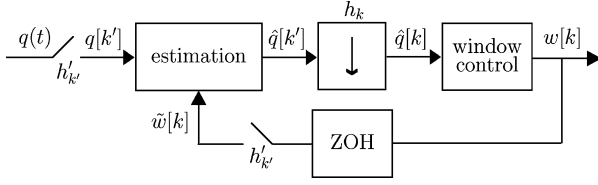


Fig. 4. Split view of the FAST TCP window update mechanism.

The estimate  $\hat{q}_n[k]$  is formed from queueing delay samples measured at each acknowledgment arrival,  $w_n[k]$  times for the  $k$ th update of (5). Denote discrete time at this timescale by  $k'$ , with inter-packet time  $h'_{k'}$ . With the obvious abuse of notation, we refer to values at this sampling time as  $q[k']$  etc.. The estimator is

$$\hat{q}_n[k' + 1] = (1 - \sigma_n[k']) \hat{q}_n[k'] + \sigma_n[k'] q_n[k'] \quad (6a)$$

$$\sigma_n[k'] = \min \left\{ \frac{\kappa_n}{w_n[k']}, \nu \right\}. \quad (6b)$$

This non-linear filter has the characteristic of a low-pass filter, with a dynamic time constant of  $\tau_n/\kappa_n$ . In the current implementation the filter parameters are  $\kappa = 3$  and  $\nu = 1/4$  [39].

A split view of the window dynamics is given in Fig. 4.

The window size,  $w_n[k]$ , is used by the transmission control, which decides the source sending rate based on  $w_n[k]$  and the rate of the received acknowledgments. Note that the transmission control is included in the (self-clocking) link model [14]. No anti-aliasing filtering is performed prior to the sampling, and the down-sampling in the FAST TCP implementation [39]. The zero-order hold (ZOH) block is a function defined by

$$f(t) = f(t_k), \quad t_k \leq t < t_{k+1}. \quad (7)$$

A sampler with sampling instants  $t_k$  is described by

$$f[k] = \int_0^\infty \delta(s - t_k) f(s) ds, \quad k = 0, 1, 2, \dots \quad (8)$$

where  $\delta(\cdot)$  is a Dirac impulse. The two sampling rates present in the system are related by

$$h_k = \sum_{k'=1}^{w_n[k]} h'_{k'+k'_k} \quad (9)$$

where  $k'_k = \sum_{i=1}^{k-1} w_n[i]$  is the value of  $k'$  just before the start of the  $k$ th RTT. Note that due to the  $w_n[k]$  in (9), the relation between the two sampling times is time varying.

For the linear stability analysis, the dynamics (5) and (6) are modeled as

$$G_{w_n q}(s) = -\frac{\gamma_n \alpha_n d_n / q}{s \tau_n^2 + \gamma_n q} \cdot \frac{1}{s \tau_n / K_n + 1}, \quad K_n \geq 1 \quad (10)$$

valid for small perturbations around the equilibrium. Here,  $K_n$  reflects both  $\kappa_n$  and the sampling processes. This is derived in Appendix C. The first part corresponds to the window update

mechanism; while the second filter models the phase-loss due to the estimation procedure and the various holds in the system.

Finally, the backward transport delays are modeled around equilibrium as

$$q_n(t) = p(t - \tau_n^b) \quad (11)$$

where  $\tau_n^b$  is the equilibrium backward queueing delay.

### B. Loop Gain

Combining the source dynamics (10) and the Laplace transform of the communicated corrupted price (queueing delay) (11) with the queue dynamics transfer function (4), results in a negative feedback system with open loop transfer function

$$L(s) = \sum_{n=1}^N G_{p w_n}(s) G_{w_n q}(s) = \sum_{n=1}^N \mu_n L_n(s) \mu_n \quad (12a)$$

where

$$\mu_n = \frac{\alpha_n}{c q} = \frac{\alpha_n}{\sum_m \alpha_m} \quad (12b)$$

$$L_n(s) = \frac{s + \frac{1}{\tau_n}}{s + \frac{1}{\hat{\tau}}} \frac{(\tau_n - q) \gamma_n e^{-\tau_n s}}{\tau_n^2 s + \gamma_n q} \frac{1}{\frac{s \tau_n}{K_n} + 1} \quad (12c)$$

$$\frac{1}{\hat{\tau}} = \sum_{n=1}^N \mu_n \frac{1}{\tau_n}. \quad (12d)$$

Note that  $\alpha_n$  is the number of packets a source tries to queue in the network, thus the queueing delay  $q = 1/c \sum_n \alpha_n$  in equilibrium, and hence the last inequality in (12b).

*Remark 1:* We will sometimes let  $q \rightarrow 0$ ; it is then assumed that  $\alpha_n \rightarrow 0$  with  $\alpha_m/\alpha_n$  fixed, so that  $\mu_n$  is well defined. ■

When no cross traffic is present, i.e.,  $x_c = 0$ , we can interpret  $\hat{\tau}$  as a weighted harmonic mean value of the round trip delays  $\tau_n$ . In particular, when all flows have equal  $\alpha_n$ , giving  $\mu_n = 1/N$ ,  $\hat{\tau}$  is the harmonic mean of  $\tau_n$ .

### C. Stability Analysis

The full model (12) of the dynamics of FAST with the joint link model contains details concerning the equilibrium queueing delay and the RTT estimator. As a first step toward proving stability, Lemma 1 abstracts away that detail, to give a condition on the gain parameter sufficient for stability.

*Lemma 1:* Denote the half plane under the line that passes  $-1 + j0$  with slope  $1/(\omega \hat{\tau})$  by

$$H(\omega) = \left\{ z \mid \arg(z + 1) - \arctan\left(\frac{1}{\omega \hat{\tau}}\right) \in (-\pi, 0) \right\}. \quad (13)$$

Let  $F_n(s)$ ,  $n = 1, \dots, N$ , be stable transfer functions, and

$$\frac{1}{\gamma_n} > -F_{\min, n} \equiv -\min_{\theta > 0} \operatorname{Re}(F_n(j\theta)) > 0.$$

Then a system with open loop gain

$$L(j\omega) = \sum_{n=1}^N \frac{\mu_n}{\tau_n} \frac{\gamma_n F_n(j\omega)}{j\omega + \frac{1}{\hat{\tau}}} \quad (14)$$

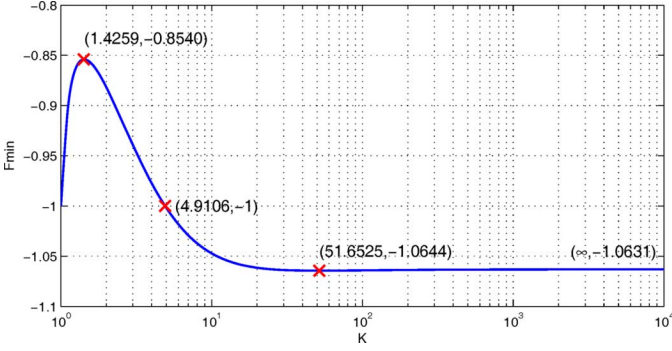


Fig. 5. Numerical solution of  $\min_{\theta} F(\theta, K)$  as a function of the filter parameter  $K \geq 1$ .

where  $1/\hat{\tau} \equiv \sum_{n=1}^N \mu_n/\tau_n$ , satisfies  $L(j\omega) \in H(\omega)$  for all  $\omega$ , and is closed loop stable for any  $\tau_n > 0, \mu_n > 0, n = 1, \dots, N$ .

*Proof:* By definition,  $L(j\omega) \in H(\omega)$  is equivalent to

$$\arg(L(j\omega) + 1) - \arctan\left(\frac{1}{\omega\hat{\tau}}\right) \in (-\pi, 0). \quad (15)$$

Substituting (14) and noting that

$$\arg\left(j\omega + \frac{1}{\hat{\tau}}\right) + \arctan\left(\frac{1}{\omega\hat{\tau}}\right) = \frac{\pi}{2} \quad (16)$$

condition (15) can be further rewritten as

$$\arg\left(\sum_{n=1}^N \frac{\mu_n}{\tau_n} \gamma_n F_n(j\omega) + j\omega + \frac{1}{\hat{\tau}}\right) \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \quad (17)$$

which is equivalent to

$$\operatorname{Re}\left(\sum_{n=1}^N \frac{\mu_n}{\tau_n} \gamma_n F_n(j\omega) + j\omega + \frac{1}{\hat{\tau}}\right) > 0. \quad (18)$$

Since  $\operatorname{Re}(\gamma_n F_n(j\omega)) > -1$  by hypothesis, and furthermore  $1/\hat{\tau} = \sum_{n=1}^N \mu_n/\tau_n$  by definition, it is established that  $L(j\omega) \in H(\omega)$  for all  $\omega$ . Thus, since

$$(-\infty, -1] \cap \bigcup_{\omega>0} H(\omega) = \emptyset$$

the Nyquist curve for  $L$  cannot encircle  $-1$ . The stability of  $F_n(s)$  implies that  $L(j\omega)$  is open loop stable, and hence the system is closed loop stable by the Nyquist criterion. ■

The construction used for Lemma 1 is depicted in Fig. 6, for  $\tau_1 = 1, \tau_2 = 5, \mu_1 = \mu_2 = 1/2$ , at  $\omega\hat{\tau} = 3$ .

*Remark 2:* The techniques used here are significantly different from ones in the existing literature on linear stability of TCP, in two respects. First, the usual approach is to find a convex hull that contains all individual  $L_n(j\omega)$  curves and then argue that any convex combination of them is still contained by the convex hull. See for example [35], [24], [6]. However, the proof of Lemma 1 deals directly with  $L(j\omega)$  instead of  $L_n(j\omega)$ . Second, for each  $\omega$ , a separate region is found to bound  $L(j\omega)$  away from the interval  $(-\infty, -1]$ . That is, the half plane  $H(\omega)$  defined by (13) depends on  $\omega$ . In existing works, convex regions are typically used to bound the whole curves and hence are independent of  $\omega$ . One exception is [26], where the frequency range

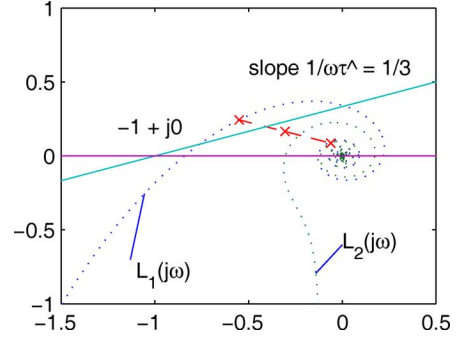


Fig. 6. An example of a line of slope  $1/(\omega\hat{\tau})$  which bounds  $L(j\omega)$ , denoted by the center cross. Note that the individual terms  $L_n(j\omega)$ , denoted by the individual crosses, are not all below this line.

is divided into two parts and different convex regions are used in the two parts. These two features lead to tighter bounds, which is necessary for the analysis of this problem. ■

In the FAST model (12c), the case when the queueing delay is  $q = 0$  is intuitively the least stable, as increasing  $q$  reduces the gain and introduces phase lead, both of which intuitively improve stability. To formalize this, Lemma 2 will be used to place bounds on the values of  $F_{\min,n}$  used in Lemma 1.

*Lemma 2:* Consider a complex half-plane  $H = \{z : \operatorname{Im}((z - z^*)e^{j\beta}) \geq 0\}$  containing 0. Consider also a function  $G : \mathbb{R}^+ \mapsto \mathbb{C}$  with  $G(\nu) = r(\nu)e^{j\theta(\nu)}$ , where  $r \geq 0$  and  $\theta$  are continuous decreasing real functions,  $r$  is unbounded as  $\nu \rightarrow 0$ , and  $\beta + \theta(0) = \Phi \in (0, \pi)$  is the angle between the edge of the half-plane and the tail of the spiral  $G$ . If  $\{G(\nu) : \nu \in \mathbb{R}^+\} \subset H$  then  $\{\eta(\nu)e^{j\phi(\nu)}G(\nu) : \nu \in \mathbb{R}^+\} \subset H$  for any  $\eta : \mathbb{R}^+ \mapsto [0, 1]$ ,  $\phi : \mathbb{R}^+ \mapsto [0, \pi - \Phi]$ .

In particular, if  $\theta(0) = -\pi/2$  then taking  $\beta = \pi/2$  gives

$$\min_{\nu} \operatorname{Re}(\eta(\nu)e^{j\phi(\nu)}G(\nu)) \geq \min_{\nu} \operatorname{Re}(G(\nu)) \quad (19)$$

for any  $\eta : \mathbb{R}^+ \mapsto [0, 1]$ ,  $\phi : \mathbb{R}^+ \mapsto [0, \pi]$ .

*Proof:* First, consider the tail of the spiral,  $\nu \in [0, \bar{\nu}]$  where  $\theta(0) - \theta(\bar{\nu}) = \phi$ . The image of the tail (under the rotation and scaling) is entirely in the sector  $\{z : \arg(z) \in [\theta(0), \theta(0) + \phi]\}$ , which is entirely in  $H$  by the definitions of  $\phi$  and  $\Phi$ .

The next step is to show that the image of any point with  $\nu > \bar{\nu}$  is also within  $H$ . Since  $H$  is convex and  $0 \in H$ , it suffices to show that for any  $\nu, \eta r(\nu) \leq r(\psi)$  where  $\theta(\psi) = \theta(\nu) + \phi$ . (Note that  $\psi > 0$  since  $\nu > \bar{\nu}$ .) But  $\psi < \nu$  as  $\theta$  is decreasing, whence  $r(\nu) \leq r(\psi)$  since  $r$  is decreasing. The result follows since  $\eta \leq 1$ .

The special case (19) follows when  $\beta = \pi/2, \Phi = 0$ . ■

It is now possible to show that FAST is stable in single-bottleneck networks.

*Theorem 3:* If for all  $n$ ,  $\gamma_n$  satisfies

$$0 < \gamma_n < \frac{1}{\sqrt{1 + \left(\frac{6}{5\pi}\right)^2}} \approx 0.93 \quad (20)$$

then a system with loop gain  $L(s)$  given by (12) is stable for arbitrary  $\alpha_n > 1, K_n \geq 1, 0 \leq q < \tau_n$ , for all  $n = 1, \dots, N$ , and arbitrary  $c > 0$ .

*Proof:* Let

$$G_K(\theta) = \frac{e^{-j\theta}}{j\theta} \frac{1+j\theta}{1+\frac{j\theta}{K}} \quad \text{and} \quad G_{\tau,K}(\theta) = G_K(\theta) \frac{\tau - q}{\tau + \frac{\gamma q \tau}{j\theta}}$$

so that (12) becomes

$$L(j\omega) = \sum_{n=1}^N \frac{\mu_n \gamma_n}{\tau_n} \frac{G_{\tau_n, K_n}(\omega \tau_n)}{j\omega + \hat{\tau}}.$$

Further, let

$$F(\theta, K) = \frac{1 - \frac{1}{K}}{1 + \frac{\theta^2}{K^2}} \cos(\theta) - \frac{1 + \frac{\theta^2}{K}}{1 + \frac{\theta^2}{K^2}} \frac{\sin(\theta)}{\theta}. \quad (21)$$

Noting that  $\text{Re}(G_K(\theta)) = F(\theta, K)$ , Lemma 4 in Appendix D implies

$$G_{\min} \equiv \min_{\theta \geq 0, K \geq 1} \text{Re}(G_K(\theta)) \geq -\sqrt{1 + \left(\frac{6}{5\pi}\right)^2}.$$

By Lemma 2,  $G_{\min} \leq G_{\min, n} \equiv \min_{\theta \geq 0} \text{Re}(G_{\tau_n, K_n}(\theta))$ . By (20),

$$\frac{1}{\gamma_n} \geq -G_{\min} \geq -G_{\min, n} > 0 \quad (22)$$

and the result follows from Lemma 1 with  $F_n(j\omega) = G_{\tau_n, K_n}(\omega \tau_n)$ . ■

*Remark 3:* Note that an effective value of  $\gamma = 0.5$  is used for the FAST implementation [39]. For this case, Theorem 3 immediately establishes FAST's stability for any pattern of round trip delays. This explains why FAST has been stable for all experimental cases studied. ■

Theorem 3 does not require  $\gamma$ ,  $\alpha$  or  $K$  to be equal for all flows. Flows may adjust  $0 < \gamma < 0.93$  and  $K \geq 1$  independently, for example using smaller  $K$  if their estimates of  $q$  are more noisy or smaller  $\gamma$  if smooth rate changes are preferable to a rapid response.

Numerical solution of  $\min_{\theta} F(\theta, K)$  as a function of  $K$ , plotted in Fig. 5 shows several interesting features. The bound in (20) has 0.6% slackness. The minimum value is approximately  $F(2.6809, 51.6525) \approx -1.0644$ , while the right hand side of (20) is approximately  $1/1.0705$ . The bound will now be further relaxed for several interesting cases of  $K$ .

For  $K \in [1, 4.9]$ ,  $\min_{\theta} F(\theta, K) > -1$ . Thus, the system is stable for all  $\gamma \in (0, 1]$ . The implementation of FAST uses  $\kappa = 3$ . Taking the dynamics of the estimator and the hold functions into account, this can be reasonably approximated by (10) with  $K \approx 1/(1/\kappa + 1/2) = 6/5$ . This suggests that FAST TCP is stable for  $\gamma \in (0, 1]$ , as stated in [39].

Ignoring filtering and phase loss due to hold functions in the system,  $K \rightarrow \infty$ . Then the bound on  $\gamma$  can be relaxed marginally to  $-1/F_{\min} \approx 1/1.0639 \approx 0.94$ , where  $F_{\min} \equiv \min_{\theta} F(\theta, \infty)$ . This agrees with the earlier analysis of this case in [33]. More importantly, the simplification  $K \rightarrow \infty$  allows further insight into the impact on stability of the distribution of round trip times in the network.

Theorem 3 is proved by finding a uniform bound for all flows'  $\tau_n$ . If we have more detailed knowledge about the round trip

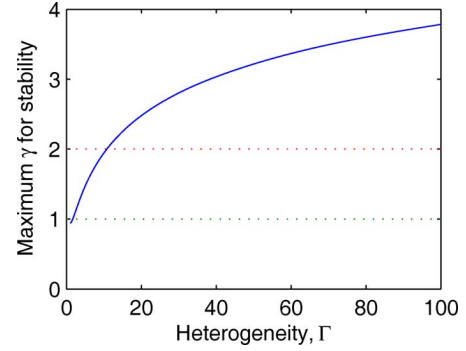


Fig. 7. Maximum value for  $\gamma$  for stability with RTTs uniform in  $(1, \Gamma)$ .

delay distribution, we may achieve tighter bounds. If (18) holds for all  $\omega$ , then (14) is closed-loop stable. When  $q \ll \tau$  and  $K \gg 1$ , the window mechanism is much slower than the rest of the source dynamics, and is hence dominant. When  $q \rightarrow 0$  (see Remark 1) and  $K \rightarrow \infty$ , then (18) holds if

$$\sum_{n=1}^N \frac{\mu_n \gamma_n}{\tau_n} \left( \cos(\omega \tau_n) - \frac{\sin(\omega \tau_n)}{\omega \tau_n} \right) + \frac{1}{\hat{\tau}} > 0. \quad (23)$$

This was studied in detail in [33]. Let us now explore it for some special cases.

1)  $N = 1$ : If there is a single flow,  $\mu_1 = 1$  in (12) and the joint link model degenerate to the static link model. In this case, FAST is stable for all  $\gamma < \pi/2$ . In this case, (20) and (23) are loose simply because the frequency,  $\omega = 2.7437/\tau$ , which minimizes  $F(\omega \tau, \infty)$ , does not coincide with a frequency at which the Nyquist plot of  $L(s)$  crosses the real axis.

2)  $N = 2$ : Consider two FAST flows with  $\mu_1 = \mu_2$  (corresponding to the current practice that  $\alpha$  is identical for all flows). Write  $\tau_1 = \Gamma \tau_2$ , where  $\Gamma$  measures the heterogeneity. Define

$$\gamma_{\Gamma} = \sup_{\theta} \frac{-(\Gamma + 1)}{\Gamma \left( \cos(\theta) - \frac{\sin(\theta)}{\theta} \right) + \cos(\Gamma \theta) - \frac{\sin(\Gamma \theta)}{\Gamma \theta}}. \quad (24)$$

From (23), a sufficient condition for stability of (12) is  $\gamma < \gamma_{\Gamma}$ . It was stated in [33] that  $\gamma_{\Gamma}$  increases from  $\gamma_1 = -1/F_{\min}$  to a peak of 1.294 at  $\Gamma = 2$  and then decreases towards  $-1/F_{\min}$  as  $\Gamma \rightarrow \infty$ . Closer analysis [15] shows that, while the values tabulated in [33] are largely correct,  $\gamma_{\Gamma}$  is discontinuous, and bounded above by 1 for irrational  $\Gamma$ .

3)  $N = \infty$ : In reality, the link is likely to be shared by many flows. It is then interesting to find the statistical mean value of the stability bound for those scenarios. We will now consider the case of many flows with continuously distributed RTTs, letting  $\alpha_n \rightarrow 0$  with  $\alpha_m/\alpha_n$  fixed.

Let  $M(\tau) = \sum_{\tau_n \leq \tau} \alpha_n/(cq)$ , and let all  $\tau_n$  be in the range  $\Omega = (\underline{\tau}, \bar{\tau})$ , with  $\bar{\tau}$  possibly infinite. If there are many flows with RTTs drawn from a continuous distribution, then applying  $\mu(\tau) = M'(\tau)$  to (23) gives

$$\int_{\Omega} \frac{\mu(\tau)}{\tau} \left( \cos(\omega \tau) - \frac{\sin(\omega \tau)}{\omega \tau} \right) d\tau + \frac{1}{\gamma} \int_{\Omega} \frac{\mu(\tau)}{\tau} d\tau > 0. \quad (25)$$

Noting that



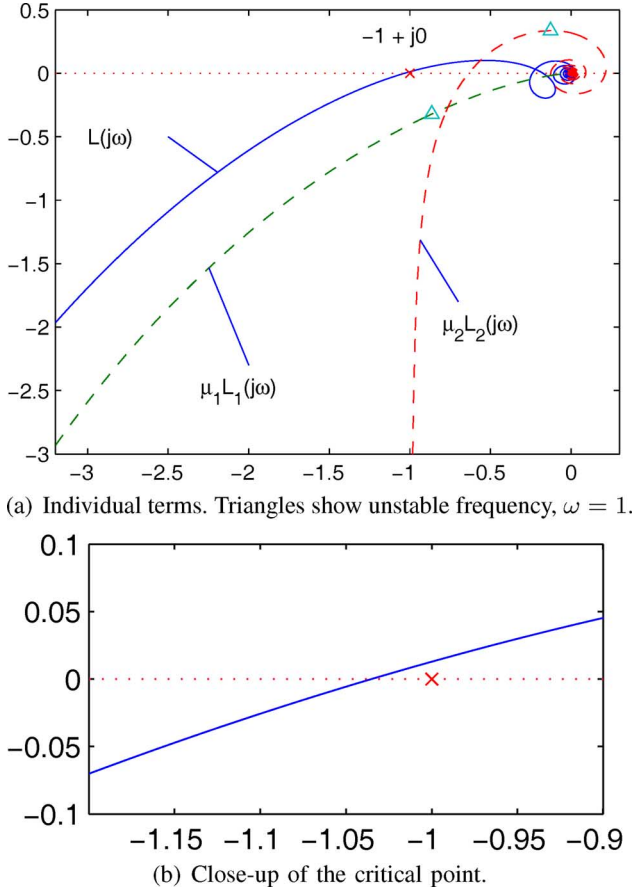


Fig. 8. Nyquist plot of the system of Section III-D.

$$\frac{d}{d\theta} \frac{\sin(\theta)}{\theta} = \frac{\cos(\theta)}{\theta} - \frac{\sin(\theta)}{\theta^2}$$

and setting  $\theta = \omega\tau$ , (25) becomes

$$\frac{1}{\gamma} > - \frac{\int_{\omega\Omega} \mu\left(\frac{\theta}{\omega}\right) \left(\frac{\sin(\theta)}{\theta}\right)' d\theta}{\int_{\omega\Omega} \frac{\mu(\theta/\omega)}{\theta} d\theta} \quad (26)$$

$$= - \frac{\left[ \mu\left(\frac{\theta}{\omega}\right) \frac{\sin(\theta)}{\theta} \right]_{\omega\Omega}^{\omega\bar{\Omega}} + \int_{\omega\Omega}^{\omega\bar{\Omega}} \frac{\mu'(\theta/\omega)}{\omega} \frac{\sin(\theta)}{\theta} d\theta}{\int_{\omega\Omega}^{\omega\bar{\Omega}} \frac{\mu(\theta/\omega)}{\theta} d\theta} \quad (27)$$

where  $\omega\Omega = (\omega\Omega, \omega\bar{\Omega})$  and  $(\cdot)'$  denotes derivative. For stability, this must hold for all  $\omega > 0$ .

As an example, assume RTTs follow a uniform distribution. As units of time are arbitrary, this can be modeled without loss of generality as

$$\mu(\tau) = \begin{cases} \frac{1}{(\Gamma-1)} & \tau \in (1, \Gamma) \\ 0 & \text{otherwise,} \end{cases} \quad (28)$$

with  $\Gamma > 1$ . In that case, (26) becomes

$$\frac{1}{\gamma} > \max_{\omega > 0} \left( - \frac{\left[ \frac{\sin(\theta)}{\theta} \right]_{\omega}^{\omega\Gamma}}{\int_{\omega}^{\omega\Gamma} \frac{d\theta}{\theta}} \right) \quad (29)$$

$$= \max_{\omega > 0} \left( \frac{\sin(\omega)\Gamma - \sin(\omega\Gamma)}{\omega\Gamma \log(\Gamma)} \right). \quad (30)$$

It is easily verified [33], that the right hand side approaches  $-F_{\min}$  as  $\Gamma \rightarrow 1$ , while for  $\Gamma > 1$  the bound is strictly looser as shown in Fig. 7, and tends to  $1/\log(\Gamma)$  for large  $\Gamma$ .

#### D. A Counter-Example

Because Theorem 3 proves stability for  $\gamma$  so close to 1, it is tempting to seek to show stability for all  $\gamma \leq 1$ . However, the following example breaks that hope. It illustrates the tightness of our result with the current model and how heterogeneity can potentially hurt stability.

1) *Example 4: A Counter Example:* Consider a network with a single bottleneck link carrying two flows. The flows have RTTs  $\tau_1 = 0.1$  ms and  $\tau_2 = 2750$  ms, with  $\alpha_1 = 1$  and  $\alpha_2 = 10^8$  with  $c = 10^{18}$  packets per second. This gives  $q = 10^{-10}$ ,  $\mu_1 = 10^{-8}$ ,  $\mu_2 = 1 - \mu_1$  and  $\hat{\tau} = 2749$  ms.<sup>4</sup>

With  $\gamma = 1$ , in contrast to the implemented  $\gamma = 0.5$ , (12) is unstable in this extreme case. Instability arises because of the high heterogeneity between the RTTs of the flows, and the greater heterogeneity between the weights given to the flows.

The Nyquist curve for this network with  $\gamma = 1$  is shown with the solid line in Fig. 8(a). The dashed lines show the individual curves  $\mu_1 L_1(j\omega)$  and  $\mu_2 L_2(j\omega)$ , and the triangles show these curves for the frequency  $\omega = 1$  at which  $L(j\omega)$  first crosses the real axis. The magnified view of this curve near the point  $-1 + j0$  in Fig. 8(b) shows that the Nyquist curve does indeed encircle  $-1$  and the resulting system is unstable.

In this example, most of the weight is given to flow 2, and the instability occurs when  $\omega\tau_2 \approx 2.75$ , minimizing (23) and giving  $F(\omega\tau_2, \infty) \approx -1.06$  (see for example Fig. 5). Although  $\tau_1$  is very small,  $\mu_1$  is even smaller, making the coefficient  $\mu_1/\tau_1$  in (23) negligible, and allowing (23) to be violated by the  $\tau_2$  term. However, the extra factor of  $1/\tau_1$  provided by the numerator of the first factor in (12c) allows the imaginary part of  $\mu_1 L_1(j)$  to balance that of  $\mu_2 L_2(j)$  where the curve crosses the axis.

This example shows that two flows are sufficient to cause instability, even though a network with a single flow (or multiple homogeneous flows) is always stable. It is also possible to construct a network of three flows with slightly less extreme parameters ( $\mu_1 = 1.2 \times 10^{-5}$ ,  $\mu_2 = 0.982$ ,  $\mu_3 = 0.0179$ ,  $\tau_1 = 300$  ms,  $\tau_2 = 150.1$  ms and  $\tau_3 = 0.2$  ms). The final Nyquist plot looks very similar to that of Fig. 8(a).

#### E. Closed Loop Validation

This subsection studies cases when  $\gamma > 1$ . Stability predictions based on the three different link models are compared with packet level simulations, complementing the open loop validation in Section II-B. This example has moderately heteroge-

<sup>4</sup>This example is rather extreme. The fluid approximation breaks down as packets are not small relative to  $\alpha$ ; in the real system,  $q$  would be dominated by packetization jitter. However, the example is of theoretical interest as it gives an upper bound on the  $\gamma$  which can yield guaranteed stability of the model, to complement the lower bound of Theorem 3. Unstable cases with realistic parameters appear in Section III-E.



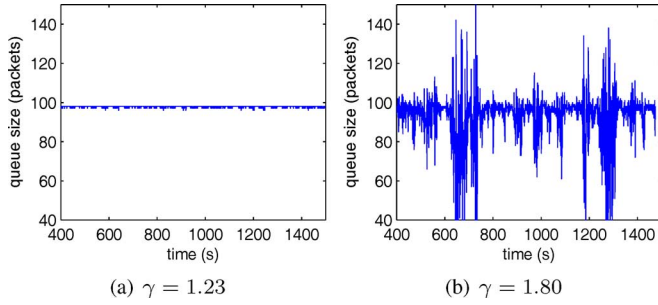


Fig. 9. Queue trajectories with critical step sizes predicted by the integrator link model and the static link model.

neous RTTs; qualitative differences occur for highly disparate RTTs, as shown in [32].

1) *Example 5: Closed Loop Validation:* Two FAST TCP flows share a single link with capacity of 10000 pkt/s. The propagation delays of the two flows are 400 ms and 700 ms. Both flows use  $\alpha = 50$  and  $K = \infty$ . The open loop transfer functions for all three models and the critical step size ( $\gamma_c$ ) for stability predicted by those models are summarized below. The integrator link model predicts a critical step size much smaller than that from the static link model, while the joint link model yields a prediction in between as expected.

- Integrator link model:  $\gamma_c = 1.23$

$$L(s) = \sum_{n=1}^N \mu_n \frac{\frac{1}{\tau_n}}{s + \frac{1}{\tau}} \frac{d_n \gamma_n e^{-\tau_n s}}{\tau_n^2 s + \gamma_n q}. \quad (31)$$

- Static link model:  $\gamma_c = 1.80$

$$L(s) = \sum_{n=1}^N \mu_n \frac{d_n \gamma_n e^{-\tau_n s}}{\tau_n^2 s + \gamma_n q}. \quad (32)$$

- Joint link model:  $\gamma_c = 1.69$

$$L(s) = \sum_{n=1}^N \mu_n \frac{s + \frac{1}{\tau_n}}{s + \frac{1}{\tau}} \frac{d_n \gamma_n e^{-\tau_n s}}{\tau_n^2 s + \gamma_n q}. \quad (33)$$

We now report NS-2 packet level simulations [7].<sup>5</sup>Fig. 9 shows the queue trajectories with  $\gamma = 1.23$  and 1.80, the critical step size to maintain stability predicted by the integrator link model and the static link model. It is clear that the queue is not stable with  $\gamma = 1.80$ , which means the static link model is too optimistic for stability analysis. We further show queue trajectories with  $\gamma = 1.65$  and 1.75 in Fig. 10. The case with  $\gamma = 1.65$  is still stable which suggests that the integrator link model is too conservative, at least in this case, while the queue starts to oscillate with  $\gamma = 1.75$ , suggesting that the critical step size is indeed approximately  $\gamma_c = 1.69$ . The fluctuations are not due to burstiness; fluctuations of the same magnitude were observed in the window sizes.

<sup>5</sup>To validate the link model, the code was modified to update the window once per RTT, and for modeling simplicity the RTT estimate was evaluated over 0.1 RTT. All queue trajectories are plotted after initial transients, to emphasize the local stability of the congestion avoidance phase.

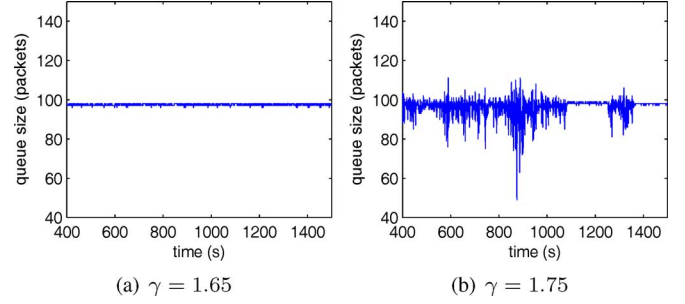


Fig. 10. Queue trajectories around critical step sizes.

#### IV. CONCLUSIONS AND FURTHER WORK

We have proposed and analyzed a link model which captures the queue dynamics when congestion windows of TCP sources change. The model is shown to be much more accurate than existing ones. It agrees with the known static link model when flows' round trip delays are similar, and approximates the standard integrator link model when the heterogeneity of round trip delays is significant. Using this new model, we have shown that FAST TCP is always linearly stable with a single bottleneck link. This extends the existing stability results on homogeneous FAST flows to cases with heterogeneous delays and resolves the notable discrepancy between empirical observations and existing theoretical predictions. The analysis highlights the critical role of self-clocking in TCP stability and the scalability of FAST TCP with respect to delay. The proof technique used here is new and less conservative than the existing ones, which is necessary for the analysis of this problem. Throughout this paper, various open loop and closed loop simulations are used to validate our predictions. In particular, we are able to predict the stability region of the closed loop system accurately compared to packet level simulations.

There are several possible directions in which to extend this work. Although the joint link model (3) improves on previous models by considering the window-based self-clocking, it still estimates flow rates by an Euler approximation over one RTT. This approximation does not become exact in the fluid limit as the inter-packet time vanishes. Even more accurate (but less tractable) continuous time models could be derived from discrete time models like (34), which is very accurate at the sampling points. Preliminary results for a single flow [15] yield a differential algebraic equation, which reduces to the static, integrator and joint link models under different orders of Padé approximation. The model remains to be extended to general networks of multiple flows and multiple links, and tractable approximations remain to be found. Another possible extension is to investigate things beyond stability, e.g., it will be of great interest to see its prediction of performance, such as queue distribution.

#### APPENDIX A SIMPLIFIED JOINT LINK MODEL DERIVATION

This section outlines the derivation of (3). For simplicity, it considers a single flow with no forward delay sending over a single bottleneck link with static non-window-based cross

traffic. For more details, including the more general multiflow case with time varying cross traffic, see [14].

Consider the evolution of the queue at time instants  $t_k$  where  $t_{k+1} = t_k + d + p_k$ . The term “window size” here denotes the actual number of outstanding packets in the network, which may in practice sometimes be different from the TCP congestion window. The amount of data from the source that arrives to the queue in the interval  $(t_k, t_{k+1}]$  corresponds to the window size at the end of that interval,  $w_{k+1}$ . During the interval,  $(d + p_k)x_c$  cross traffic will also have arrived, and  $(d + p_k)c$  data will have departed the queue. Dividing these contributions to the change in queue size by  $c$  to convert to delay, the queueing delay is updated according to

$$p_{k+1} = p_k + \frac{w_{k+1}}{c} + \frac{d + p_k}{c}(x_c - c). \quad (34)$$

Note that the sample time  $h_k \equiv d + p_k$  (i.e., round trip delays) of the discrete time dynamical system (34) is non-uniform. This model is known from [14] to be extremely accurate at its sampling instants.

To arrive at a continuous time approximation of the discrete model that includes a direct term, (34) can be formulated as

$$\frac{p_{k+1} - p_k}{h_k} = \frac{1}{c} \left( \frac{w_k}{d + p_k} + x_c - c \right) + \frac{1}{c} \frac{w_{k+1} - w_k}{h_k} \quad (35)$$

and then, by the means of a first order Euler approximation of the derivative, approximated in continuous time as

$$\dot{p}(t) = \frac{1}{c} \left( \frac{w(t)}{d + p(t)} + \dot{w}(t) + x_c - c \right). \quad (36)$$

Note that the accuracy of the derivative approximation depends on the RTT rather than the number of packets in the system (the window size). Thus this modeling error remains even in the fluid limit. Solutions to this limitation are currently under investigation. However, the joint link model is shown to be accurate by validation examples in [14], and its multiflow version (3) is shown in Examples 1–3 of Section II to be an improvement on its predecessors.

## APPENDIX B

### FAST TCP STABILITY UNDER THE INTEGRATOR MODEL

It was stated in [37] that the integrator link model predicts that for all  $\gamma$ , there is a  $\tau$  such that a single FAST flow will be unstable. This was the original motivation for proposing the joint link model (3). However, the model (1), also an integrator, actually predicts that FAST should be stable for all  $\gamma \leq 1$ . This follows from Lemma 1, as the second term in (21) disappears. Comparing this to the counter-example in Section III-D shows that the more accurate model (3) predicts FAST to be *less* stable in the worst case than a pure integrator (1) does.

The instability predicted in [37] was in fact due to the details of the integrator link model studied. There are two important differences between the model of [37] and (1): (a) an extra factor of  $\tau$  in the gain in [37], (b) an additional delay of  $\tau$  in the denominator in [37].

We will now show that the extra factor of  $\tau$  in the gain in [37] is an error. When that is fixed, the model (still with the additional

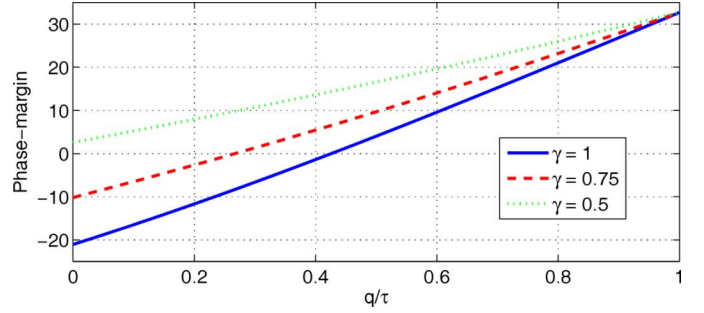


Fig. 11. Phase-margin plot.

delay of  $\tau$ ) in fact predicts that for some  $\gamma$  there is a maximum  $\tau/q$  for stability, while for others a single flow is always stable.

Modeling the link and ignoring estimator dynamics as in [37], the loop gain for the single source case is

$$\bar{L}(s) = \frac{\gamma \frac{x}{c} \frac{d}{\tau^2} \frac{e^{-s\tau}}{\tau}}{\left(s + \frac{x}{c} \frac{e^{-s\tau}}{\tau}\right) \left(s + \gamma \frac{q}{\tau^2}\right)}. \quad (37)$$

Using  $\tau = d + q$ , the poles of the closed loop system are given by the characteristic equation

$$s \left(s + \gamma \frac{q}{\tau^2}\right) + \frac{x}{c} \frac{e^{-s\tau}}{\tau} \left(s + \frac{\gamma}{\tau}\right) = 0 \Leftrightarrow 1 + \tilde{L}(s) = 0 \quad (38)$$

where

$$\tilde{L}(s) = \frac{x}{c} \frac{e^{-s\tau}}{s\tau} \frac{\left(s + \frac{\gamma}{\tau}\right)}{\left(s + \frac{\gamma q}{\tau^2}\right)} = \frac{x}{c} \frac{e^{-s\tau}}{s\tau} \frac{(s\tau + \gamma)}{\left(s\tau + \frac{\gamma q}{\tau}\right)}. \quad (39)$$

We can study the poles of the closed loop system (i.e., the zeros of the characteristic equation) via the Nyquist theorem using either the loop gain  $\bar{L}(s)$  or the transformed loop gain  $\tilde{L}(s)$ . Due to the exponential function in the denominator in  $\tilde{L}(s)$  it is more convenient to use  $\tilde{L}(s)$  in the analysis. Note that  $\tilde{L}(s)$  is similar to the loop gain studied in [37]; what differs is the extra factor of  $\tau$  used in [37], which arose from the use of a round-trip-time timescale for the window dynamics and an absolute timescale for the link dynamics, without explicit conversion. Let  $\theta = \omega\tau$ . For the case of  $\gamma = 1$ , the argument of the numerator of (39) when evaluated at  $s = j\theta$  is

$$\arg e^{-j\theta} + \arg(j\theta + 1) = -\theta + \arctan(\theta) < 0. \quad (40)$$

From (39), stability depends on the ratio between the queue delay and the round trip delay, i.e.,  $q/\tau$ . Taking  $\tau/q$  arbitrarily large gives  $\arg \tilde{L}(j0) \rightarrow -\pi$ , and since  $|\tilde{L}(j0)| \rightarrow \infty$ , the system must encircle  $-1 + j0$ , and is hence not stable by the Nyquist theorem. Fig. 11 shows the phase-margin for different ratios  $q/\tau$  for different  $\gamma$ 's. Observe that for  $\gamma = 0.5$  the model system is stable for all  $q$  and  $\tau$ .

This shows how sensitive stability is to the delay in the link model. To justify the absence of delay in the denominator of (3), note that this is the case in which the pole and zero cancel, to give exact agreement with the initial rise experimentally observed in Figs. 1–3.

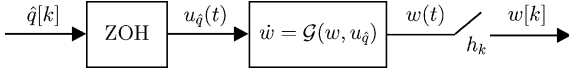


Fig. 12. Continuous time equivalent window control.

## APPENDIX C

## LINEAR CONTINUOUS TIME SOURCE DYNAMICS

Since sources work in time-scales of round trip delays, which are usually heterogeneously distributed, and there are different, time-varying sampling times within the source dynamics, we derive continuous time models that are amenable to analysis and equivalent to (5) and (6) at sampling instants. These are used to derive a continuous time approximate mapping  $q(t) \mapsto w(t)$ , valid for small perturbations.

Consider first the window control block of Fig. 4. We are seeking the nonlinear function  $\mathcal{G}(w, u_q)$  in Fig. 12. Note that (5) has the form

$$w[k+1] = \mathcal{F}(\hat{q}[k])w[k] + \gamma\alpha. \quad (41)$$

Using this insight we assign

$$\dot{w} = \mathcal{G}(w, u_q) = g(u_q)w + g_0(u_q). \quad (42)$$

Since  $g(u_q)$  is constant over the sample interval  $t_k \leq t < t_{k+1}$ , the state at  $t_{k+1}$  is given by

$$w(t_{k+1}) = \exp(g(u_q(t_k))h_k)w(t_k) + g_0(u_q) \int_0^{h_k} \exp(g(u_q(t_k))s) ds$$

where  $h_k = t_{k+1} - t_k = d + q(t_k)$  is the observed round trip delay. Identification with (41) yields

$$g(u_q) = \frac{\log(\mathcal{F}(u_q))}{h_k} = \frac{\log(1 - \gamma\xi)}{h_k} \quad (43)$$

$$g_0(u_q) = \frac{\gamma\alpha g(u_q)}{\exp(g(u_q)h_k) - 1} = -\frac{\alpha \log(1 - \gamma\xi)}{h_k \xi} \quad (44)$$

where  $\xi = u_q/(d + u_q)$ , so that (42) becomes

$$\dot{w}(t) = \frac{\log(1 - \gamma\xi(t))}{h_k(t)} \left( w(t) - \frac{\alpha}{\xi(t)} \right). \quad (45)$$

Consider now the estimator block in Fig. 4, whose split view given in Fig. 13 describes  $\mathcal{H}(\hat{q}, u_q, u_w)$ . Deriving  $\mathcal{H}$  from (6) is analogous to deriving  $\mathcal{G}$  from (5), cf. (41)–(45). Repeating this procedure, and equating to (6) instead of (5) at sampling instants, gives

$$\dot{\hat{q}}(t) = \frac{\log(1 - \sigma(t))}{h'_k(t)} (\hat{q}(t) - u_q(t)) \quad (46)$$

where  $\sigma(t) = \min\{\kappa/u_w(t), \nu\}$  and  $u_w(t)$  is the sampled-and-held  $w[k']$  (cf. (6b)).

Assuming that  $d$  is estimated accurately, the system can be linearized around an equilibrium point  $(w, q)$ . Note that, by (46), the feedback loop in Fig. 4 will not be active in the linear case since  $\hat{q} = q = u_q$  in equilibrium. In a high capacity, large latency network, we typically have  $\gamma\hat{q}/\tau \ll 1$  and  $\kappa/w \ll \nu < 1$ . Thus  $\log(1 - \gamma\xi) = \log(1 - \gamma q/\tau) \approx -\gamma q/\tau$  and  $\log(1 - \kappa/w) \approx -\kappa/w$ . Furthermore, the sampling intervals

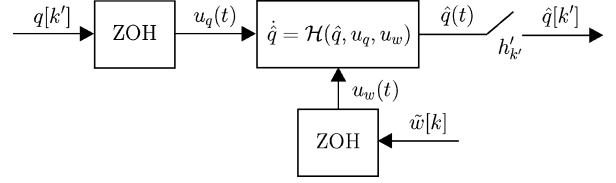


Fig. 13. Continuous time split view of the estimator dynamics.

are approximately time invariant around the equilibrium, i.e.,  $h_k(t) = \tau$  and  $h'_k(t) = \tau/w$  for each individual source. Under these assumptions,  $\alpha/q = w/(d + q) = (c - x_c)\alpha_n / \sum_m \alpha_m$ . Replacing variables by perturbations from the equilibrium, the linearized versions of (45) and (46) become

$$\dot{w}(t) = -\gamma \frac{q}{\tau^2} w(t) - \gamma \frac{\alpha d}{q\tau^2} u_q(t) \quad (47)$$

$$\dot{\hat{q}}(t) = -\frac{\kappa}{\tau} \hat{q}(t) + \frac{\kappa}{\tau} u_q(t). \quad (48)$$

Assuming constant sample intervals  $h$ , the Laplace transform of (8) is unity, while that of the zero-order hold (7) is

$$\frac{1 - e^{-sh}}{sh} \approx \frac{1}{\frac{sh}{2} + 1} \quad (49)$$

taking a Padé approximation of order (1, 1) of the exponential. Thus, under the high capacity large delay assumption, the approximate time constant of the individual zero-order holds associated with the estimator dynamics is  $\tau/(2w)$ , i.e., the zero-order hold on the delay input in Fig. 13. This is substantially faster than the time constants of the estimator dynamic (48), which is  $\tau/\kappa$ , and of the window dynamics (47), which is  $\tau^2/(\gamma q)$ . Therefore the dynamics of this zero-order hold can be ignored. However, the approximate time constant of the zero-order hold in Fig. 12 is  $\tau/2$ , which is of the same order of magnitude as the estimator dynamics time constant (note that the time scales of the two dynamics are separated when  $\kappa \gg 1$ ). In conclusion, by adding the filter

$$\frac{1}{\frac{s\tau}{K} + 1}, \quad K \geq 1 \quad (50)$$

to the source window control dynamics, with  $1/K \approx 1/\kappa + 1/2$ , the model includes phase-loss due to the queueing delay estimation and the different holds in the system.

APPENDIX D  
BOUNDS ON  $F(\theta, K)$ 

The main result of this appendix is Lemma 4, the proof of which uses two additional lemmas. For details, see [15].

*Lemma 4:* For all  $\theta \geq 0$  and  $K \geq 1$ ,

$$F(\theta, K) \geq -\sqrt{1 + \left(\frac{6}{5\pi}\right)^2}. \quad (51)$$

where  $F(\theta, K)$  is given by (21). More specifically,

$$\min_{\theta \geq 0, K \geq 1} F(\theta, K) = F^* \equiv \frac{\theta^* \cos(\theta^*) - \sin(\theta^*) - \sqrt{1 + \theta^{*2}}}{2\theta^*} \quad (52)$$

where

$$\theta^* \equiv \min \left\{ \theta > \frac{\pi}{2} \mid f_\theta(\theta) = 0 \right\} \in \left( \frac{5\pi}{6}, \pi \right). \quad (53)$$

and

$$f_\theta(\theta) = \frac{1}{\sqrt{1+\theta^2}} + (1-\theta^2)\sin(\theta) - \theta\cos(\theta). \quad (54)$$

*Proof:* The proof considers separately  $\theta$  in the intervals  $[0, \pi/4]$ ,  $[\pi/4, 5\pi/6]$  and  $(5\pi/6, \infty)$ , and applies Lemma 5.

For  $\theta \in [0, \pi/4]$ , (57) does not hold, and since  $\cos(\theta) > 0$ , Lemma 5 gives  $F(\theta, K) \geq -1$ , whence (51).

For the other two cases, note that  $|a\sin(\theta) + b\cos(\theta)| \leq \sqrt{a^2 + b^2}$  for all  $a, b$ , whence by Lemma 5, for all  $\theta > 0$ ,

$$F(\theta, K) \geq -\frac{\sqrt{1+\theta^2}}{\theta}. \quad (55)$$

For  $\theta \in (5\pi/6, \infty)$ , this establishes (51), since the right hand side is increasing in  $\theta$ .

For the remaining case, using the definition (58), Lemma 6 establishes that

$$\min_{\theta \in [\pi/4, 5\pi/6]} F_{K^*}(\theta) = F_{K^*}\left(\frac{5\pi}{6}\right) \quad (56)$$

giving (51) in all three cases.

Lemma 6 shows the ' $\in$ ' of (53). To obtain (52), note first that the value on the right hand side is achievable. For  $\theta \in [0, \pi/4]$ ,  $F(\theta, K) \geq -1$  as shown above. For  $\theta \in [\pi/4, \pi]$ , the minimum is  $F^*$  by Lemma 6. To see that  $F$  is not minimized by some  $\theta > \pi$ , observe that (55) is tight as  $K \rightarrow \infty$  for the  $\theta^+ \in (5\pi/6, \pi)$  which solves  $\theta^+ \tan(\theta^+) = -1$ . Thus  $F^* \leq -\sqrt{1+(\theta^+)^2}/\theta^+$  by (63) of Lemma 6. Since the right hand side of (55) is increasing, no  $F(\theta, K)$  can be less than  $F^*$  for  $\theta > \theta^+$ , giving (52). ■

*Lemma 5:* For all  $\theta \geq 0$  and  $K \geq 1$ , if

$$K^*(\theta) \equiv \theta \frac{\sqrt{1+\theta^2} - (\theta \cos(\theta) - \sin(\theta))}{\cos(\theta) + \theta \sin(\theta)} \in [1, \infty) \quad (57)$$

then

$$\min_{K \geq 1} F(\theta, K) = F_{K^*}(\theta) \equiv \frac{\theta \cos(\theta) - \sin(\theta) - \sqrt{1+\theta^2}}{2\theta}. \quad (58)$$

otherwise,

$$\inf_{K \geq 1} F(\theta, K) = \min(\cos(\theta), 0) - \frac{\sin(\theta)}{\theta} \quad (59)$$

while the right hand side of (58) is a lower bound.

*Proof:* First consider the stationary points of  $F(\theta, K)$  with respect to  $K$ . Any stationary point  $K^+(\theta)$  satisfies

$$0 = \frac{\partial F}{\partial K} = \frac{\theta^2(2K-1) + K^2}{(\theta^2 + K^2)^2} \cos(\theta) - \frac{\theta^2(\theta^2 - K(K-2)) \sin(\theta)}{(\theta^2 + K^2)^2} \frac{\sin(\theta)}{\theta}.$$

This gives the only stationary extrema at

$$K^+(\theta) = \pm \theta \frac{\sqrt{1+\theta^2} - (\theta \cos(\theta) - \sin(\theta))}{\cos(\theta) + \theta \sin(\theta)} \quad (60)$$

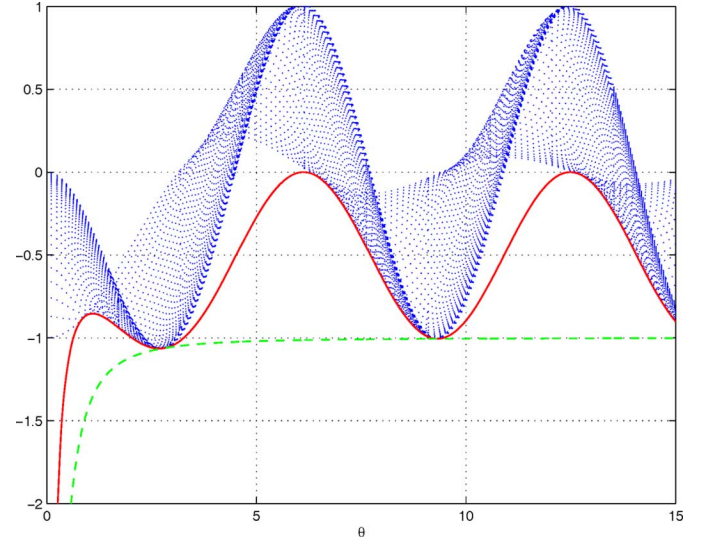


Fig. 14. Dotted lines:  $F(\theta, K)$  for logarithmically spaced  $K \in [1, 1000]$ . Solid line:  $F_{K^*}(\theta)$ . Dashed line:  $-\sqrt{1+\theta^2}/\theta$ .

and consequently

$$F_{K^+}(\theta) = \mp \frac{\sqrt{1+\theta^2} - \theta \cos(\theta) + \sin(\theta)}{2\theta}. \quad (61)$$

Noting that  $|\theta \cos(\theta) - \sin(\theta)| \leq \sqrt{1+\theta^2}$ , the ' $-$ ' of the  $\mp$  yields a negative value for any  $\theta > 0$  while the ' $+$ ' yields a positive value. Thus, if the infimum of  $F(\theta, K)$  is at a stationary point, it is at  $K^*$ .

The infimum will be either at a stationary point or for  $K$  at the boundary of the feasible region  $[1, \infty)$ . Clearly  $F(\theta, 1) = \sin(\theta)/\theta > F(\theta, K^*)$  and  $\lim_{K \rightarrow \infty} F(\theta, K) = \cos(\theta) - \sin(\theta)/\theta > F(\theta, K^*)$ . This establishes that  $F_{K^*}(\theta)$  is a lower bound.

If  $K^*(\theta) \in [1, \infty)$  then  $F_{K^*}(\theta)$  is feasible, and (58) is established. Otherwise, the infimum is on the boundary, and (59) follows from (21), taking the infimum rather than the minimum since the feasible region is half open. ■

The result of Lemma 5 is illustrated in Fig. 14.

*Lemma 6:* For any  $\phi \in [\pi/2, \theta^*]$ , with  $\theta^*$  given by (53),

$$\min_{\theta \in [\pi/4, \phi]} F_{K^*}(\theta) = F_{K^*}(\phi). \quad (62)$$

Moreover,  $\theta^* \in (5\pi/6, \pi)$  and

$$\min_{\theta \in [\pi/4, \pi]} F_{K^*}(\theta) = F_{K^*}(\theta^*). \quad (63)$$

*Proof:* First, consider the interval  $\theta \in [\pi/4, \pi/2]$ . Now  $F_{K^*}(\pi/4) > F_{K^*}(\pi/2)$ , and furthermore

$$\left. \frac{dF_{K^*}}{d\theta} \right|_{\theta=\pi/4} = f_\theta\left(\frac{\pi}{4}\right) > 0. \quad (64)$$

This implies that  $F_{K^*}(\theta)$  has at least one extremum in the considered interval corresponding to a local maximum. Hence  $f_\theta(\theta) = 0$  has at least one solution in the same domain. But since, for all  $\theta \in [\pi/4, \pi/2]$ ,

$$\frac{df_\theta}{d\theta} = -\theta \frac{\sqrt{1+\theta^2}}{(1+\theta^2)^2} - \theta \sin(\theta) - \theta^2 \cos(\theta) < 0$$

the solution must be unique, so

$$\min_{\theta \in [\pi/4, \pi/2]} F_{K^*}(\theta) = F_{K^*}\left(\frac{\pi}{2}\right). \quad (65)$$

Now, consider the interval  $\theta \in [\pi/2, \phi]$ . To establish (62), it is now sufficient to show that

$$f_\theta \leq 0 \text{ for all } \theta \in \left[\frac{\pi}{2}, \phi\right]. \quad (66)$$

First, note that  $\theta^* \in (\pi/2, \pi)$  because  $f_\theta(\pi/2) < 0$  and  $f_\theta(\pi) > 0$ . Since  $f_\theta(\pi/2) < 0$  and  $f_\theta(\theta^*) = 0$  we only need to check if there exist any positive local maxima. Any extremum is given by solving

$$-\frac{1}{\theta} \frac{df_\theta}{d\theta} = (1+\theta^2)^{-3/2} + \sin(\theta) + \theta \cos(\theta) = 0. \quad (67)$$

Now

$$\frac{d}{d\theta} \left( -\frac{1}{\theta} \frac{df_\theta}{d\theta} \right) = -3\theta(1+\theta^2)^{-3/2} + 2\cos(\theta) - \theta \sin(\theta) < 0 \quad (68)$$

over  $[\pi/2, \pi] \supset [\pi/2, \theta^*]$ . Thus (67) has at most one feasible solution. But

$$\left. \frac{df_\theta}{d\theta} \right|_{\theta=\pi/2} = -\frac{\pi}{2} \left( \left( 1 + \left( \frac{\pi}{2} \right)^2 \right)^{-3/2} + 1 \right) < 0 \quad (69)$$

implies that this stationary point is a minimum, after which  $f_\theta$  is increasing. Thus

$$\max_{\theta \in [\pi/2, \theta^*]} f_\theta(\theta) = \max \left\{ f_\theta\left(\frac{\pi}{2}\right), f_\theta(\theta^*) \right\} = 0 \quad (70)$$

and thus (62).

By (68) and (69),  $f_\theta$  is increasing on  $(\theta^*, \pi]$ , but  $f_\theta(\theta^*) = 0$ , whence  $f_\theta(\theta) > 0$  for  $\theta > \theta^*$ . This establishes (63). Moreover,  $f_\theta(5\pi/6) < 0$ , whence  $\theta^* > 5\pi/6$ . ■

#### ACKNOWLEDGMENT

The authors thank D. Wei, F. Paganini, and K. H. Johansson for valuable discussions.

#### REFERENCES

- [1] T. Alpcan and T. Basar, "Global stability analysis of an end-to-end congestion control scheme for general topology networks with delay," in *Proc. IEEE Conf. Decision Control*, 2003, pp. 1092–1097.
- [2] E. Altman, C. Barakat, and V. Ramos, "Analysis of AIMD protocols over paths with variable delay," in *Proc. IEEE INFOCOM*, Mar. 2004, pp. 993–1000.
- [3] L. L. H. Andrew, S. V. Hanly, and R. G. Mukhtar, "CLAMP: A system to enhance the performance of wireless access networks," in *Proc. IEEE Globecom*, 2003, pp. 4142–4147.
- [4] F. Baccelli and D. Hong, "AIMD, fairness and fractal scaling of TCP traffic," in *Proc. IEEE INFOCOM*, 2002, pp. 229–238.
- [5] L. S. Brakmo and L. L. Peterson, "TCP Vegas: End-to-end congestion avoidance on a global internet," *IEEE J. Selected Areas Commun.*, vol. 13, no. 8, pp. 1465–1480, Oct. 1995.
- [6] H. Choe and S. H. Low, "Stabilized Vegas," in *Proc. IEEE INFOCOM*, Apr. 2003, pp. 2290–2300.
- [7] T. Cui and L. Andrew, FAST TCP Simulator Module for NS-2 Tech. Rep. [Online]. Available: <http://www.cubinlab.ee.mu.oz.au/ns2fasttcp>

- [8] J. Choi, K. Koo, J. Lee, and S. H. Low, "Global stability of FAST TCP in single-link single-source network," in *Proc. IEEE Conf. Decision Control*, 2005, pp. 1837–1841.
- [9] S. Deb and R. Srikant, "Global stability of congestion controllers for the internet," *IEEE Trans. Automat. Control*, vol. 48, no. 6, pp. 1055–1060, Jun. 2003.
- [10] W. Gong and T. Basar, "Guest editorial for the special issue on systems and control methods for communication networks," *IEEE Trans. Automat. Control*, vol. 47, no. 6, pp. 877–879, Jun. 2002.
- [11] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A Survey of recent results in networked control systems," *Proc. IEEE*, vol. 95, no. 1, pp. 138–162, Jan. 2007.
- [12] C. Hollot and Y. Chait, "Nonlinear stability analysis for a class of TCP/AQM schemes," in *Proc. IEEE Conf. Decision Control*, 2001.
- [13] C. Hollot, V. Misra, D. Towsley, and W. Gong, "A control theoretic analysis of RED," in *Proc. IEEE INFOCOM*, 2001, pp. 1510–1519.
- [14] K. Jacobsson, H. Hjalmarsson, and N. Möller, "ACK-clock dynamics in network congestion control – An inner feedback loop with implications on inelastic flow impact," in *Proc. IEEE Conf. Decision Control*, San Diego, CA, Dec. 2006, pp. 1882–1887.
- [15] K. Jacobsson, L. L. H. Andrew, A. Tang, S. H. Low, and H. Hjalmarsson, An Improved Link Model for TCP-Like Flow Control and Its Application to FAST TCP Tech. Rep. [Online]. Available: <http://netlab.caltech.edu/publications/LinkModelTR07.pdf>
- [16] R. Johari and D. Tan, "End-to-end congestion control for the Internet: Delays and stability," *IEEE/ACM Trans. Networking*, vol. 9, no. 6, pp. 818–832, Dec. 2001.
- [17] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, pp. 237–252, 1998.
- [18] R. King, R. Baraniuk, and R. Riedi, "TCP-Africa: An adaptive and fair rapid increase rule for scalable TCP," in *Proc. IEEE INFOCOM*, 2005, pp. 1838–1848.
- [19] K. Kaneko, T. Fujikawa, Z. Su, and J. Katto, "TCP-fusion: A hybrid congestion control algorithm for high-speed networks," in *Proc. PFLDnet*, 2007, pp. 31–36.
- [20] K. Kim, A. Tang, and S. H. Low, "A stabilizing AQM based on virtual queue dynamics in supporting TCP with arbitrary delays," in *Proc. IEEE Conf. Decision Control*, 2003, pp. 3665–3670.
- [21] S. Liu, T. Basar, and R. Srikant, "Pitfalls in the fluid modeling of RTT variations in window-based congestion control," in *Proc. IEEE INFOCOM*, 2005, pp. 1002–1012.
- [22] Y. Liu, F. L. Presti, V. Misra, D. Towsley, and Y. Gu, "Fluid models and solutions for large-scale IP networks," in *Proc. ACM SIGMETRICS*, 2003, pp. 91–101.
- [23] S. H. Low, F. Paganini, and J. C. Doyle, "Internet congestion control," *IEEE Control Syst. Mag.*, vol. 22, no. 1, pp. 28–43, Feb. 2002.
- [24] S. H. Low, F. Paganini, J. Wang, and J. C. Doyle, "Linear stability of TCP/RED and a scalable control," *Comput. Networks J.*, vol. 43, no. 5, pp. 633–647, 2003.
- [25] L. Massoulié, "Stability of distributed congestion control with heterogeneous feedback delays," *IEEE Trans. Automat. Control*, vol. 47, no. 6, pp. 895–902, Jun. 2002.
- [26] F. Paganini, Z. Wang, J. C. Doyle, and S. H. Low, "Congestion control for high performance, stability and fairness in general networks," *IEEE/ACM Trans. Networking*, vol. 13, pp. 43–56, Feb. 2005.
- [27] G. Raina, "Local bifurcation analysis of some dual congestion control algorithms," *IEEE Trans. Automat. Control*, vol. 50, no. 8, pp. 1135–1146, Aug. 2005.
- [28] H. Shimonishi, T. Hama, and T. Murase, "TCP-adaptive Reno for improving efficiency-friendliness tradeoffs of TCP congestion control algorithm," in *Proc. PFLDnet*, 2006, [CD ROM].
- [29] R. Shorten, F. Wirth, and D. Leith, "Modelling TCP in droptail and other environments," *Automatica*, vol. 43, no. 3, pp. 441–449, Mar. 2007.
- [30] R. Srikant, *The Mathematics of Internet Congestion Control*. Boston, MA: Birkhauser, 2004.
- [31] K. Tan, J. Song, Q. Zhang, and M. Sridharan, "A compound TCP approach for high-speed and long distance networks," in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006, pp. 1–12.
- [32] A. Tang, L. L. H. Andrew, K. Jacobsson, K. Johansson, S. H. Low, and H. Hjalmarsson, "Window flow control: Macroscopic properties from microscopic factors," in *Proc. IEEE INFOCOM*, Apr. 2008, pp. 91–95.
- [33] A. Tang, K. Jacobsson, L. L. H. Andrew, and S. H. Low, "An accurate link model and its application to stability analysis of FAST TCP," in *Proc. IEEE INFOCOM*, May 2007, pp. 161–169.



- [34] A. Tang, X. Wei, S. H. Low, and M. Chiang, "Heterogeneous congestion control: Efficiency, fairness and design," in *Proc. IEEE Int. Conf. Network Protocols*, Nov. 2006, pp. 127–136.
- [35] G. Vinnicombe, "On the stability of networks operating TCP-like congestion control," in *Proc. 15th IFAC World Congress Automat. Control*, Barcelona, Spain, Jul. 2002, [CD ROM].
- [36] J. Wang, A. Tang, and S. H. Low, "Local stability of FAST TCP," in *Proc. IEEE Conf. Decision Control*, Dec. 2004, pp. 1023–1028.
- [37] J. Wang, D. X. Wei, and S. H. Low, "Modeling and stability of FAST TCP," in *IMA Volumes in Mathematics and its Applications*, P. Agrawal, M. Andrews, P. J. Fleming, G. Yin, and L. Zhang, Eds. New York: Springer Science, 2006, vol. 143.
- [38] Z. Wang and F. Paganini, "Global stability with time-delay in network congestion control," in *Proc. IEEE Conf. Decision Control*, Dec. 2002, pp. 3632–3637.
- [39] D. Wei, C. Jin, S. H. Low, and S. Hegde, "FAST TCP: Motivation, architecture, algorithms, performance," *IEEE/ACM Trans. Networking*, vol. 14, pp. 1246–1259, Dec. 2006.
- [40] D. Wei, "Microscopic Behavior of Internet Congestion Control," Ph.D. dissertation, California Institute of Technology, Pasadena, 2007.
- [41] L. Ying, G. Dullerud, and R. Srikant, "Global stability of internet congestion controllers with heterogeneous delays," *IEEE/ACM Trans. Networking*, vol. 14, no. 3, pp. 579–591, Jun. 2006.



**Krister Jacobsson** received the M.S. degree in vehicle engineering and the Ph.D. degree from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2002 and 2008, respectively.

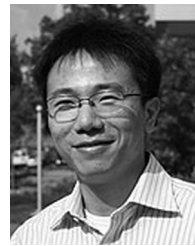
He is currently a Postdoctoral Fellow at the California Institute of Technology, Pasadena. His research interests include modeling and control of telecommunication systems.



**Lachlan L. H. Andrew** (M'97–SM'05) received the B.Sc., B.E., and Ph.D. degrees from the University of Melbourne, Melbourne, Australia, in 1992, 1993, and 1997, respectively.

Since 2008, he has been an Associate Professor at Swinburne University of Technology, Australia. From 2005 to 2008, he was a Senior Research Engineer in the Department of Computer Science, California Institute of Technology (Caltech), Pasadena. Prior to that, he was a Senior Research Fellow at the University of Melbourne and a Lecturer at RMIT, Australia. His research interests include performance analysis of congestion control, resource allocation algorithms and energy-efficient networking.

Dr. Andrew received the Best Paper Award at IEEE MASS'07. He is a member of the ACM.



**Ao (Kevin) Tang** (S'01–M'07) received the B.E. (with honor) and M.E. degrees in electronics engineering from Tsinghua University, Beijing, China, in 1999 and 2001, respectively, and the Ph.D. degree in electrical engineering (with a minor in applied and computational mathematics) from the California Institute of Technology, Pasadena, in 2006.

He is currently an Assistant Professor in the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, where his main research interests include communication networks, control and dynamical systems, optimization and game theory.

Dr. Tang received the 2006 George B. Dantzig Best Dissertation Award from INFORMS and the 2007 Charles Wilts Best Dissertation Prize from EE, Caltech.



**Steven H. Low** (M'92–SM'99–F'08) received the B.S. degree from Cornell University, Ithaca, NY, and the Ph.D. degree from the University of California at Berkeley.

He is a Professor with the Computer Science and Electrical Engineering Departments, California Institute of Technology (Caltech), Pasadena and an Adjunct Professor with the Swinburne University of Technology, Australia. He was on the Editorial Board of the *Computer Networks Journal* from 2003 to 2005. He is on the editorial boards of *ACM Computing Surveys*, *NOW Foundations*, and *Trends in Networking*. He is a Co-Editor of the Springer Book Series *Optimization and Control of Communication Systems: Theory and Applications*.

Dr. Low received the IEEE William R. Bennett Prize Paper Award in 1997 and the 1996 R&D 100 Award. He was on the editorial board of IEEE/ACM TRANSACTIONS ON NETWORKING from 1997 to 2006. He is a Senior Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is a member of the Networking and Information Technology Technical Advisory Group for the U.S. Presidents Council of Advisors on Science and Technology (PCAST).



**Håkan Hjalmarsson** was born in 1962. He received the M.S. degree in electrical engineering and the Licentiate degree and the Ph.D. degree in automatic control from Linköping University, Linköping, Sweden, in 1988, 1990, and 1993, respectively.

He has held visiting research positions at the California Institute of Technology, Pasadena, Louvain University, and the University of Newcastle, Newcastle, Australia. He is a Professor at the School of Electrical Engineering, KTH, Stockholm, Sweden.

He has served as an Associate Editor for *Automatica* (1996–2001) been a Guest Editor for the *European Journal of Control and Control Engineering Practice*. His research interests include system identification, signal processing, control and estimation in communication networks, and automated tuning of controllers.

Dr. Hjalmarsson received the KTH Award for outstanding contribution to undergraduate education in 2001. He is Vice-Chair of the IFAC Technical Committee on Modeling, Identification and Signal Processing. He has served as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL (2005–2007).